Large Scale Simulations of Sky Surveys

Abstract

Large-volume sky surveys have accessed the vast temporal and spatial expanse of the Universe via a remarkable set of measurements, and many more are sure to follow. To make new predictions for these cosmological observations, and to properly interpret them, large-scale numerical simulation and modeling has become an essential tool. Here we discuss HACC (Hardware/Hybrid Accelerated Cosmology Code), an extreme-scale N-body cosmology code and its associated analysis framework, focusing on the complexity of the analysis workflow, which is as important as running the underlying simulation.

Keywords: Supercomputing, physics

Authors: Katrin Heitmann, Salman Habib, Hal Finkel, Nicholas Frontiere, Adrian Pope, Vitali Morozov, Steve Rangel, Eve Kovacs, Juliana Kwan, Nan Li, Silvio Rizzi, Joe Insley, Venkat Vishwanath, Tom Peterka (Argonne National Laboratory), David Daniel, Patricia Fasel (Los Alamos National Laboratory), George Zagaris (Kitware)

1. Simulating the Universe as Seen through Large-Scale Sky Surveys

Cosmologists have looked deeply at the Universe and found it to be 'dark'. Detailed observations over the last three decades spanning the full range of the electromagnetic spectrum, from gamma rays to the radio sky, carried out from the ground and from space, persuasively suggest an astonishing picture: ~70% of the matter-energy content of the Universe is made up of a mysterious 'dark energy' component, potentially responsible for the Universe's accelerated expansion, 25% of the matter exists in the form of a yet unidentified 'dark matter' component, and only 0.4% of the remaining ordinary matter, happens to be visible. Understanding the physics of the mysterious dark sector is the foremost challenge in cosmology today. Major cosmological missions are ongoing and planned to create ever more detailed maps of how mass is distributed in the Universe. These maps hold the key to advancing our knowledge of the make-up and evolution of the Universe, enabling us to unlock its 'dark' secrets.

Unlike a science based on the experimental method, cosmology lacks investigations under the researcher's control, performed under strict isolation, and allowing step-by-step progress towards solving a physical problem, however complex it may be. The task is instead to make a number of robust observations, where statistical and systematic errors can be bounded, and then to arrive at scientifically defensible inferences about the Universe. To do this, one creates model Universes allowing for different cosmological models and astrophysical effects, mimicking possible observational systematics and even implementing the "clean-up" of observational data from unwanted foregrounds obscuring the signals being searched for. The

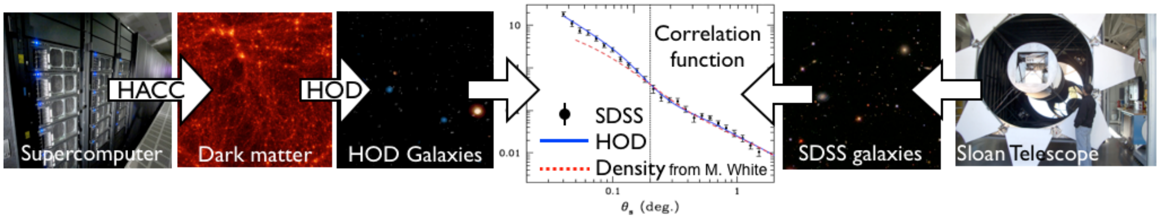


Figure 1: Pipeline to extract cosmological information from galaxy surveys. The halo occupancy distribution (HOD) is a statistical method used to 'paint' galaxies onto the dark matter distribution. The mass distribution from large simulations is populated with galaxies that live in dark matter clumps called halos, the galaxy count and brightness being correlated with the halo mass. The results are compared to the galaxy distribution as measured by cosmological surveys.

complexity of this task leads inexorably to the use of the world's largest supercomputers.

This requires an end-to-end computational approach, starting from the fundamental theory (Einstein's general relativity and modifications thereof, quantum mechanics), to simulating the formation of large-scale structures in the Universe and creating synthetic maps for the observation of interest, to modeling the instrument, and effects that can bias observations, such as atmospheric turbulence (for a detailed discussion of an end-to-end pipeline for the Large Synoptic Survey Telescope see Ref. [1]). We must understand the observed system as a whole, and determine which physics will have important effects on what scales, where simple modeling suffices, as compared to fully self-consistent simulations, and how uncertainties in modeling and simulations can bias the conclusions. This task is complicated further by the fact that we often cannot directly observe what we desire to study but must draw conclusions from an indirect analysis, as for instance in the use of baryonic tracers (galaxies) of the large-scale structure of the Universe.

In this paper, we focus on describing our current efforts to create synthetic galaxy maps from large-scale simulations for optical surveys, setting aside telescope modeling as a separate problem. In order to build these synthetic maps, we have to simulate the evolution of the mass distribution in large cosmological volumes with exquisite resolution. To demonstrate the scale of this challenge, a quick summary of the relevant scales is as follows. Modern survey depths require covering simulation volumes of order tens of cubic Gpc (1 pc=3.26 light-years); to follow bright galaxies, structures with a minimum mass of 10^{11} M¤ (M¤=1 solar mass) must be tracked and resolved by at least a hundred simulation particles. The force resolution must be small compared to the size of the objects to be resolved, i.e., ~kpc. This immediately implies a dynamic range (ratio of smallest resolved scale to box size) of a part in 10^6 (~Gpc/kpc) everywhere in the *entire* simulation volume. In terms of the number of simulation particles required, the implied counts range from hundreds of billions to many trillions. This requires access to very large supercomputers and, in today's landscape of diverse supercomputing architectures, a highly scalable and portable N-body code. To this end we have developed HACC (Hardware/Hybrid Accelerated

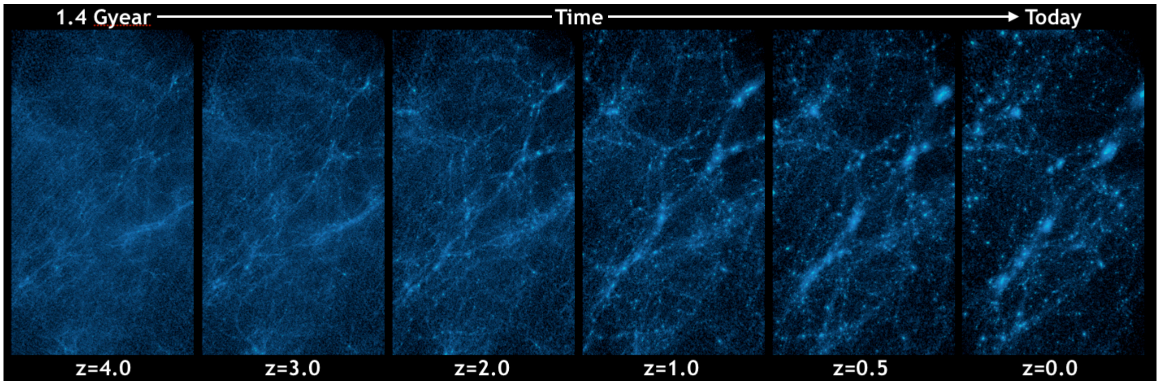


Figure 2: Time evolution of structure formation. A zoom-in into a dense region is shown. The frames depict the structure at different redshifts or times, starting 1.4 Gyears after the Big Bang. Images here and in Fig. 6 were generated using the vl3 parallel volume rendering system [2].

Cosmology Code), a high performance code framework targeted at current and future architectures. We will provide a description of HACC in Section 2.

The mass distribution in the Universe is probed indirectly as most of the mass is dark, neither emitting nor absorbing light. Since the presence or absence of light is what we observe, the connection between mass and light is of fundamental importance. To investigate this connection, a major analysis suite has to be created and a seamless workflow must exist to ingest the raw simulation output and produce large-scale maps of galaxies. We have created such an analysis environment, which combines in situ analysis tools with a suite of post-processing steps, described in Section 3. As an example, Figure 1 summarizes the analysis path for the statistics of galaxy surveys. In the article we will focus on describing the different steps on the left side of the image, how to use large-scale supercomputers to create detail maps of our Universe as seen through optical telescopes. The last section will show some concrete examples of our effort.

2. HACC: Enabling Large-Scale Structure Simulations on Modern Supercomputing Platforms

HACC was initially designed for the Roadrunner supercomputer [3,4], the first system to break the Petaflop barrier. With its novel architecture of acceleration via the Cell processor, Roadrunner was by far the most forward looking machine of its generation, providing a glimpse of the current frontier and some illumination of the path to the exascale [5].

The design of a modern high-performance code must begin with an awareness that methods and algorithms should not be developed without an understanding of future programming paradigms and computing and storage architectures. The HACC computational strategy is based on a hybrid representation of physical information on computational grids as well as ‘particles’ that, depending on the context, can be viewed as tracers of mass, or as micro-fluid elements. This hybrid representation is flexible and can be made to map well to machine architectures as well as to be aligned with multiple programming paradigms. Additionally, it provides a broad choice of methods that can be optimized given architectural, power, and other constraints, and the best combination can be picked for any given platform.

Technically speaking, HACC simulates cosmic structure formation by solving the gravitational Vlasov-Poison equation in an expanding Universe [6]. The simulation starts from a smooth Gaussian random field that evolves into a 'cosmic web' comprised of sheets, filaments, and mass concentrations called halos. An image of the formation of cosmic structures over time is given in Figure 2. The Vlasov-Poison equation is hopeless to solve as a PDE because of its high dimensionality and the development of nonlinear structure, therefore N-body methods are employed.

HACC uses a combination of grid and particle methods, where the grid methods are used to resolve the large to medium (smooth) length scales and the particle methods are employed to resolve the smaller scales. This split between the long- and short-range solver offers a very convenient organization of the code: the long-range force (in this case an FFT-based solver) exists at the higher level of the code and is essentially architecture-independent. It is implemented in C/C++/MPI and its performance and scaling is dominated by the FFT implementation. We have developed a new pencil-decomposed FFT and demonstrated scaling up to 1,572,864 cores in Sequoia, a 96-rack IBM BG/Q system [7]. The particle-based short-range solver exists at a lower level of the computational hierarchy and is architecture-tunable. It combines MPI with a variety of local programming models (OpenCL, OpenMP, CUDA) to readily adapt to different platforms. To enhance its flexibility, the short-range solver uses a range of algorithms, direct particle-particle interactions, i.e., a P^3M algorithm [8], as on Roadrunner and Titan, or both tree and particle-particle methods as on the IBM BG/Q ('PPTreePM'). The grid is responsible for 4 orders of magnitude of dynamic range, while the particle methods handle the critical 2 orders of magnitude at the shortest scales where particle clustering is maximal and the bulk of the time-stepping computation takes place. An in-depth description of the HACC design and implementation, including the long-range solver, the different short-range solvers, the time stepper, and spatial decomposition of the code, as well as its scaling properties, is given in Ref. [9].

HACC's multi-algorithmic structure attacks several weaknesses of conventional particle codes including limited vectorization, indirection, complex data structures, lack of threading, and short interaction lists. Currently, HACC is implemented on conventional and Cell/GPU-accelerated clusters [3,4,9], on the Blue Gene architecture [7], and is running on prototype Intel Xeon Phi hardware. HACC is the first, and currently the only large-scale cosmology code suite worldwide, that can run at scale on *all* available supercomputer architectures. HACC achieved outstanding performance on both Sequoia and Titan, reaching almost 14 PFlops

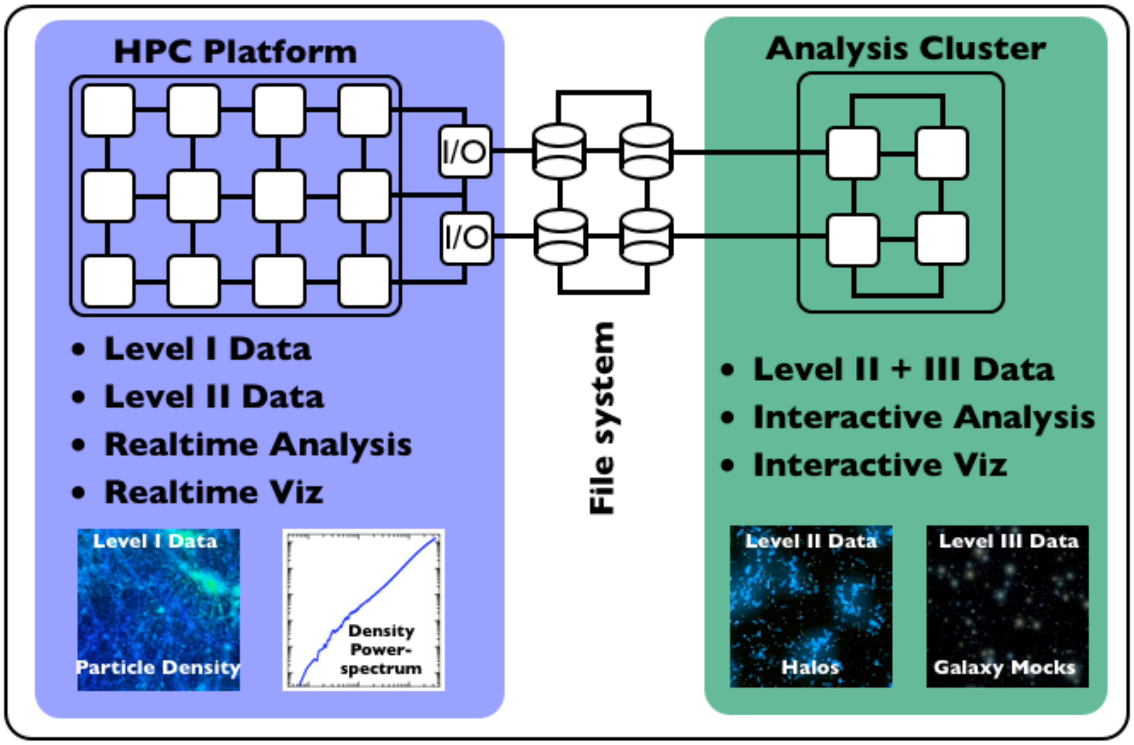


Figure 3: Data levels and analysis hierarchy of a cosmological simulation.

(69.2% of peak) on Sequoia, a kernel peak of 20.54 PFlops on 77% of Titan (the full machine was not available for these scaling runs), and 7.9 PFlops of sustained performance on 77% of Titan for the full code. HACC's outstanding performance and portability has enabled us to carry out some of the challenging simulations needed to advance our understanding of the dark Universe.

3 Analytics Requirements and Tools

Analyzing the data from the large cosmological simulations is as demanding as carrying out the simulations themselves. In fact some of the computational modeling applied to the outputs in post-processing is even more complex than the N-body simulation itself with respect to the physical processes involved. To put the challenge posed by the analysis task into context, a single time snapshot from one of the simulations discussed in Section 4 encompasses 40TB of raw data, and of the order of 100 snapshots have to be analyzed. This amount of data clearly demands a well-thought out analysis strategy, combining in situ and post-processing tools. Another challenge is posed by the fact that the raw data from the simulation is very science-rich. Not only can we generate optical catalogs from the simulation -- the example discussed in the next section -- but also field maps, e.g., the cosmic microwave background temperature, or X-ray flux. It is important to store enough of the already processed data to ensure that new science projects can be carried out at later stages.

In order to design an efficient workflow to tackle these challenges and to decide which analysis tools have to be run in situ and therefore on the HPC system itself (which means that they should scale as well as the main code, a difficult task in and of itself), it is useful to break up the data into three more or less distinct levels: (i) Level I, the raw simulation output, where the particles, densities, etc. live; (ii) Level II, the 'science' level, that is, the output rendered as a description useful for further theoretical analysis, including halo and sub-halo information, merger trees, line-of-sight skewers; and (iii) Level III, the 'galaxy catalog' level where the data is further reduced to the point that it can be interacted with in real-time. Very roughly speaking, at each higher level, the data size reduces from the previous level by 2 or 3 orders of magnitude.

The data layer plays a crucial role for science applications. Because of the imbalances in the I/O bandwidth relative to peak performance for the computation and the extreme stressing of file systems, it has been apparent for some time that dumping raw data to a storage system for post-analysis is a poor strategy for a problem where intensive analysis of very large datasets is essential. Therefore, we carry out as much of the analysis as possible on the raw Level I data on the HPC system itself, as well as the reduction of Level I data to Level II. The Level II datasets can then be loaded into an analysis cluster and further analyzed. A schematic of the different data levels and analysis hierarchy is shown in Figure 3.

Level I analysis requires algorithms for tasks such as halo-finding, determining correlation functions and a host of other statistical measures, building halo merger trees, and carrying out automated sub-sampling of the data. The overall data hierarchy must take the needs of the analysis routines as well as that of the simulation code into account, in order to maintain locality and avoid data movement. Level II data products can be used for science directly or used to produce Level III data products such as mock survey catalogs that include galaxies with realistic colors, luminosities, and morphologies. The computational algorithms we apply to address our science goals include density estimation, anomaly detection, tracking, high-dimensional model fitting, and non-parametric inversion. These techniques are computation and memory intensive and have been developed to work within the raw Level I and Level II data products. To show two concrete examples that will be important for our analysis in the next Section, we now discuss the halo finder, which runs in situ with the simulations and reduces data from Level I to Level II, and the halo merger tree code that acts on Level II data and enables the generation of Level III data.

Halo Finding: The 'halo' concept plays a very important role within cosmological simulations. Dark matter halos are the hosts of galaxies and by mapping out galaxies we can draw conclusions about the dark matter distribution in the Universe. Halos mark over-densities in the dark matter distribution and can be identified through different algorithms. Most commonly, they are found by either locating density peaks directly and growing spheres out to a characteristic over-density or via

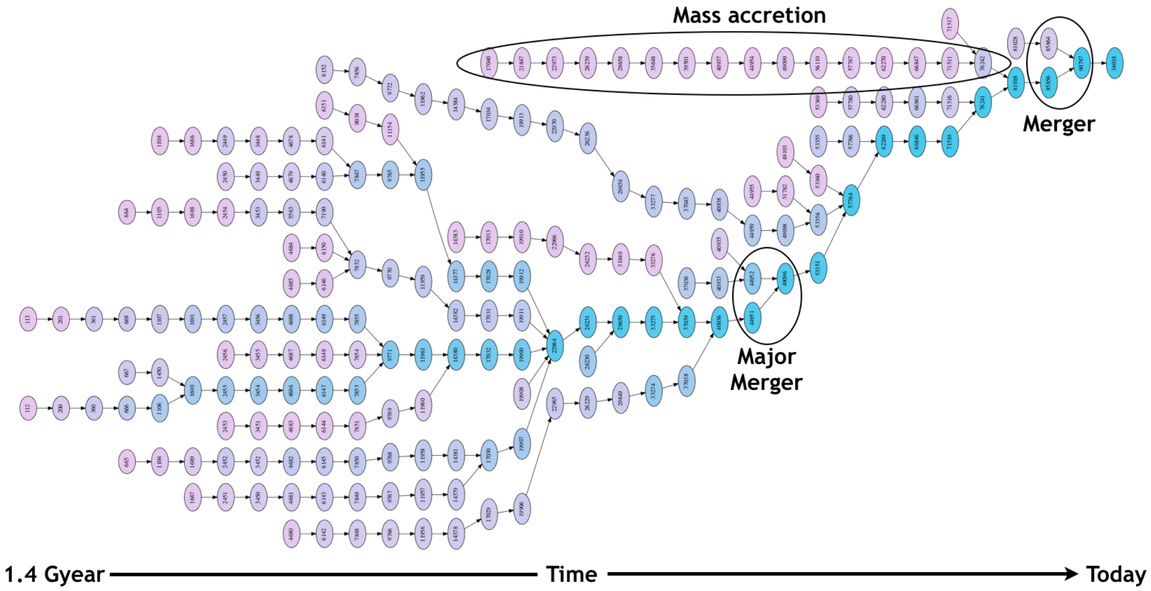


Figure 4: Merger tree for the formation of an individual halo. Each vertex in the tree shows a dark matter halo at a certain time step (time advances from left to right, vertices on each vertical line are halos that exist at the same time). Light colors depict light halos, darker, blue colors, more massive halos. Halos grow over time via two main mechanisms: (i) incremental mass accretion, (ii) merging of halos where a merger of halos with similar masses is called a "major merger". The merger tree shown here is relatively small, trees with up to 10,000 nodes can easily exist in the simulations.

neighbor finding algorithms. Here we will discuss one of the simplest, the so-called friends-of-friends (FOF) algorithm which is used for all of following results. In FOF halo finding, for each particle, every particle within a certain distance, the so-called linking length (usually between 0.15 to 0.2 of the mean inter-particle spacing) is identified as a 'friend'. The process is then continued for each friend particle, and so on. If the number of particles in such a conglomerate is above a certain threshold (usually approximately ~100 particles) the structure is called a halo. Its center is found by either finding the particles with the most friends (maximum local density) or by determining the potential minimum of the halo, or by finding the average position from all particles in the halo (the center of mass). Naively, the FOF algorithm requires N^2 operations, but the algorithm is straightforwardly sped up to NlogN via a tree implementation. In addition, our FOF finder takes full advantage of the already existing overloaded data structure strategy of the main code in order to enable parallel halo finding. Halos can be identified independently on each rank and halos on the edge of a rank are not missed since particle information is available from the neighboring ranks. A final reconciliation step ensures that halos are not counted more than once. For details of the implementation and scaling properties of the algorithm, see Ref. [10].

As mentioned previously, the FOF finder reduces the raw Level I simulation data to Level II data. The halo catalog itself, which contains information about halo properties such as position and velocities, is negligible in size compared to the raw data. In addition to the halo catalog, we store the tags of all particles in halos (depending on the threshold of what defines a halo, the number of particles in halos is approximately 50% of all particles) and their halo tag (identifying to which halo each particle belongs), which we need to construct halo merger trees, as discussed next. Finally, we store full particle information (positions and velocities) for a subset of particles in halos (usually 1%) to enable placements of galaxies at those positions later on, and all particles in halos above a large mass cut-off. This set of data (halo information and reduced information about particles in halos) defines the set of Level II data connected to the halos and reduces the data volume by a factor of approximately 10. Most of the data is stored in the particle tags of particles that are in halos -- once the halo merger trees are built, this information can be discarded and the data reduction then reaches more than a factor of 100, as stated earlier. Halo finding is carried out for roughly 20% of all global time steps (there are no halos very early in the simulation). Compared to the time stepper itself, the relative cost of the halo finder decreases over time, but is always at a comparable level of time consumption as a single time step. Because of the data size and computation time consumed, it is not feasible to offload this step to a smaller analysis cluster.

Merger Tree Construction: The FOF algorithm identifies halos from individual snapshots based on the spatial relationship between particles at a fixed point in time. In order to determine halo temporal evolution, we evaluate the FOF output from the complete sequence of snapshots. Our algorithm compares halos from adjacent snapshots, and constructs a graph for representing evolutionary events. The graph, called a merger tree, represents each halo by a vertex (see Figure 4 for an example), and similar halos in adjacent snapshots with an edge. We define a similarity measure as the fraction of shared particles (i.e., the particle intersection of two halos) to total particles from the earlier of the two halos.

To construct the merger trees between subsequently taken snapshots, it is sufficient to compare the particle membership functions obtained by the FOF finder. However, computing the pairwise similarity matrix for the halos of all the adjacent snapshots requires some efficiency. We implement a technique for determining the intersection cardinality of multiple sets that is linear with respect to the number of particles, after an initial particle sort is performed. We utilize a type of sparse matrix representation for the similarity matrix to reduce the otherwise large memory requirement. The memory reduction is significant due to the large amount of sparsity inherent to the problem, and we have witnessed the computational overhead incurred to be marginal.

As mentioned, this approach relies on the result of the FOF algorithm for calculating the halo membership function. Because the clustering algorithm requires halos to have a minimum number of particles, the hard threshold can cause some misidentification of events when halos are near the minimum cutoff value. In order to reduce these misidentifications, we maintain a windowed history of missing halos. The particles from the missing halos are stored for comparison with later snapshots to determine if they re-emerge, and if so, to be treated as coming from some pre-existing halo.

4. The Simulated and the Real Universe

Finally, we show some results from the analysis of recent simulations carried out on Mira at the Argonne Leadership Computing Facility and on Titan at the Oak Ridge Leadership Computing Facility. As mentioned in the Introduction, one important task in the analysis is to transform the mass distribution we obtain from the N-body simulations into actual galaxy catalogs. Simulating galaxies from first principles in a cosmological volume is still far from possible -- the dynamical range is vast and the physics of galaxy formation, inadequately understood. Instead, galaxies can be painted onto the dark matter distributions, using models of different levels of sophistication. The main assumption here is that "light traces mass", meaning that the galaxies trace the density of the mass distribution, which is predominantly dark. This assumption is true only as an approximation; the aim is to develop more complex prescriptions to 'light up' the dark matter distribution with galaxies.

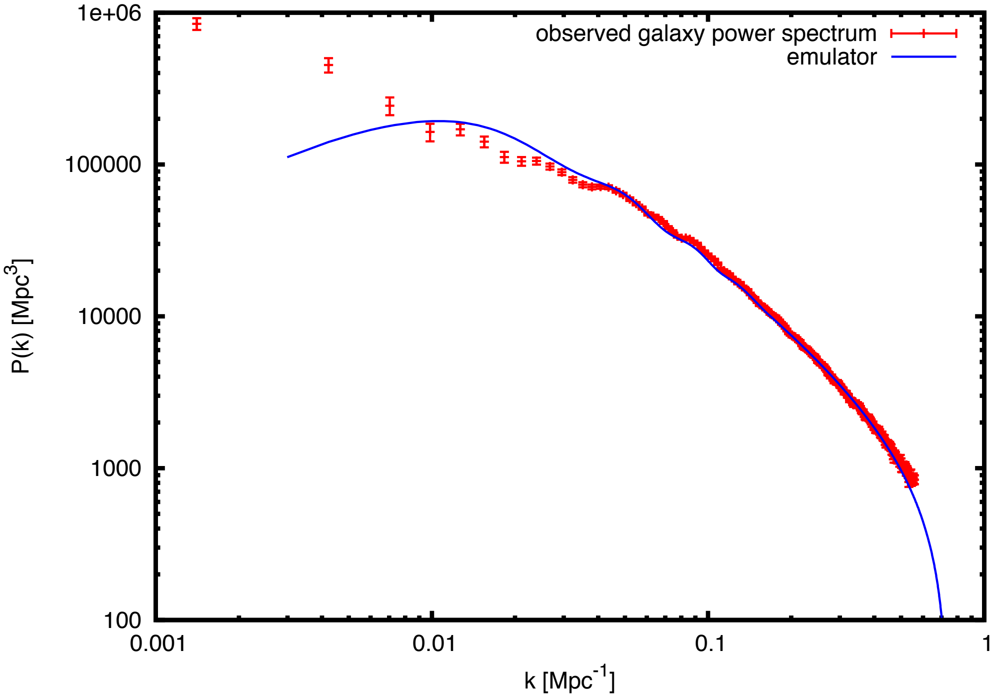


Figure 5: Comparison of a simulated galaxy spectrum to observations from the SDSS-BOSS survey [12]. The power spectrum is the Fourier analog of the two-point correlation function and characterizes the tendency of galaxies to cluster together.

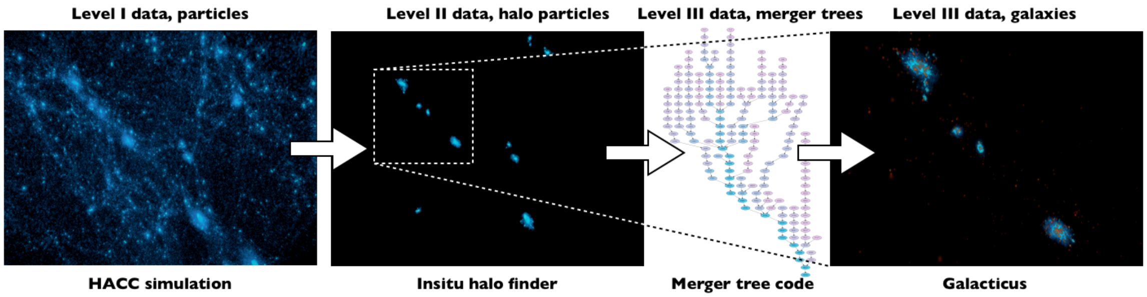


Figure 6: From the raw simulation to the galaxy catalog: the left panel shows the zoom-in to a full particle distribution from the N-body simulation (Level I data), the second panel shows the dark matter halos identified with the FOF halo finder (Level II data), the third panel shows a merger tree (Level III data), and the right panel shows the galaxies embedded in the halos as determined by Galacticus (Level III data).

A simple and powerful approach to this problem uses the so-called Halo Occupation Distribution (HOD) model [10]. In this approach, a number of so-called central and satellite galaxies of a certain type are assigned to a dark matter halo depending on the halo's mass. The central galaxy lives at the center of the halo and is the brightest galaxy. If the halo is heavy enough to host more galaxies, satellite galaxies are assigned and placed within the halo. The HOD model is described by approximately five parameters that are tuned to match one observable, e.g. the galaxy power spectrum. Once the model is fixed, other observables can be predicted from the galaxy catalog. Recently, we have built synthetic sky maps based on a large Mira simulation evolving 32 billion particles in a (2.1 Gpc)^3 volume and investigated the dependence of the galaxy power spectrum on the five HOD modeling parameters [13]. The best-fit HOD model on top of data from BOSS, the Baryon Oscillation Spectroscopic Survey, is shown in Figure 5. This figure demonstrates how well the results from simulations match the observational data.

While the HOD approach is simple, it has one major shortcoming: it completely neglects the formation history of a halo that surely will carry information about the galaxy population it hosts today. For example, if a halo formed very early and mainly grew through mass accretion, it will not have much star formation today. Or, if the halo underwent a violent merger with another large halo, it will also have a distinct galaxy population. In order to take these effects into account, so-called semi-analytic models (SAMs) have been developed. SAMs follow the evolution of each halo via halo merger trees and solve along the way a set of physics equations that describe galaxy formation in an approximate way. SAMs deliver very detailed descriptions of the galaxies that populate halos, including their colors, positions, and shapes, star formation history, black hole content, etc. The drawback of the SAMs is that they depend on a large number of parameters (two to three hundred) that have to be tuned to observations. In Figure 6 we show an example of our full simulation and analysis pipeline working to create a synthetic galaxy map. The simulation, carried out on Titan, has a mass resolution of ~10^9 M¤ and therefore can capture the smaller halos that host bright galaxies reliably. As the simulation was run, halos were

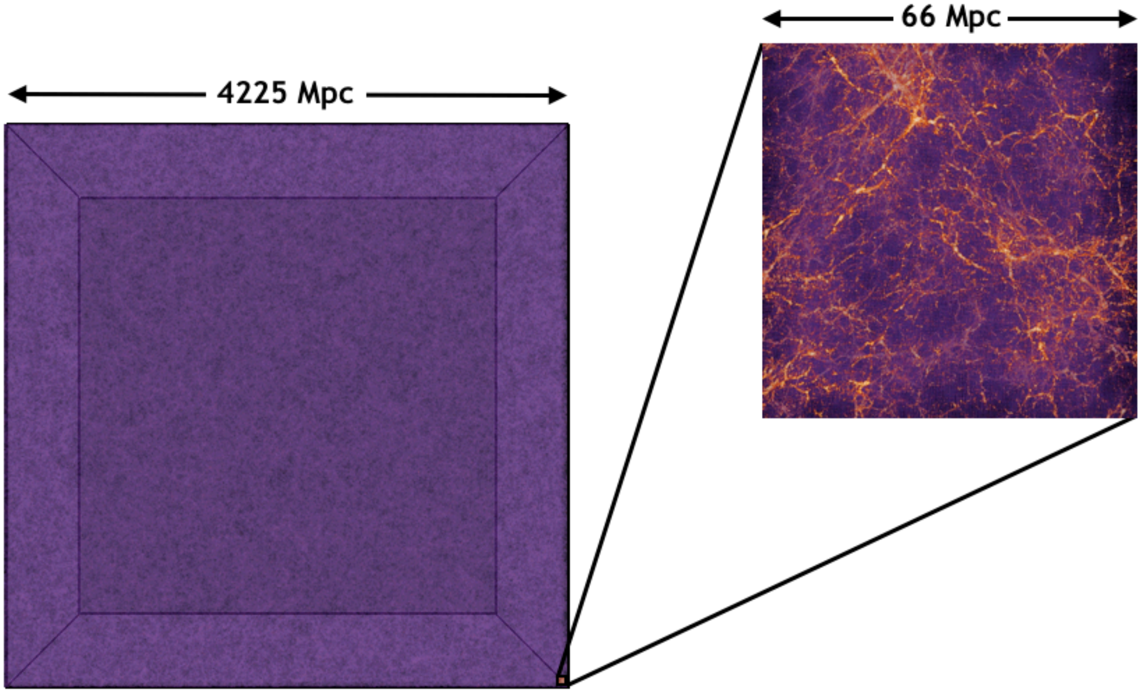
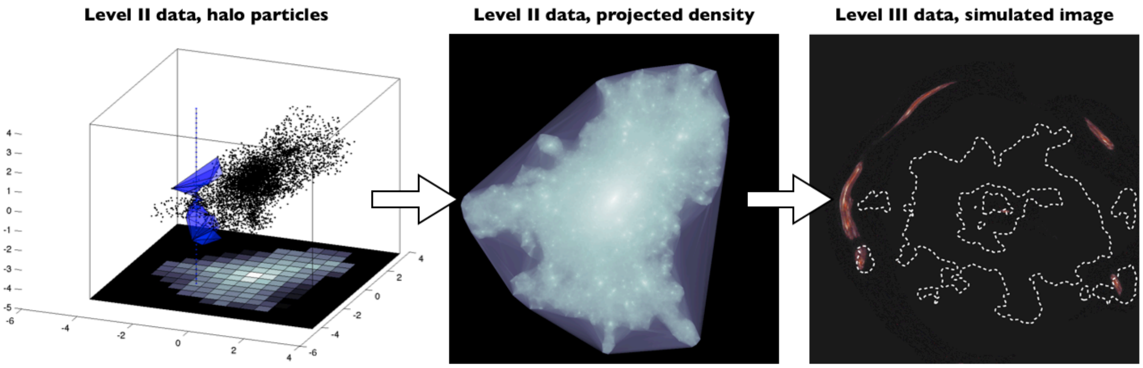


Figure 7: Dynamic range of the Outer Rim simulation: the left panel shows the full simulation volume of (4225 Mpc)^3, the right panel the output from just one of the 262,144 cores.

identified on the fly, and the information of particles resident in halos was stored. From this information, merger trees were constructed to track the evolution of each halo in detail. Finally, a sophisticated semi-analytic model was run on the merger trees, in this case Galacticus [14], to generate a full synthetic galaxy sky.

Our last example shows results from the largest cosmological simulation ever attempted: the Outer Rim simulation. This simulation is currently running on Mira at ALCF and evolves 1.1 trillion particles in a (4225 Mpc)^3 volume. As for the Titan run, each particle has a mass of ~10^9 M¤ but the volume covered is many times larger. The force resolution in the simulation is ~4.1 kpc, achieved via a 10240^3 PM mesh on the large scales in combination with the tree solver on small scales. To demonstrate the scale of this simulation, we show a slice of the full simulation box in Figure 7 as well as the output from one of the 262,144 ranks the simulation is run on. New science results have been already extracted from the simulation (even though it has not quite yet reached the present epoch) and Figure 8 shows an example of the exciting science results that can be obtained. In this case, a halo at a certain time was extracted, a tessellation-based estimator was used to create its two-dimensional projected density, and a ray-tracing code used to generate a strong gravitational lensing image from a simulated source (see Figure 8 for the workflow). Strong lensing refers to the severe distortion of galaxy images and the generation of multiple images due to the presence of a massive intervening object between the source galaxies and the observer. In our case, the halo from the simulation is the massive object (in the center of the right panel), galaxies are placed behind this lens and the visible arcs are the distorted images as given by a fast ray-tracing algorithm.



Level II data, halo particles

Level II data, projected density

Level III data, simulated image

Figure 8: From the raw simulation to a simulated strong lensing image: The left panel shows the tessellation approach for density estimation applied to the particle data extracted with the halo finder. The 2-d density field is obtained by a weighted sum of 3-d density estimates. The box size is on the scale of an individual halo (in units of Mpc), and the grid resolution is independent of simulation parameters. Points are sampled at discrete intervals on lines normal to the 2-d grid cells (blue line). Sample points within the tetrahedra intersected by the line are identified and interpolated. This way a 2-d density field is created as seen by an observer (if dark matter would be directly visible, density shown in the middle panel). Finally, galaxies are placed behind the halo and lensed images of these galaxies are created through a ray-tracing algorithm (right panel).

The resulting images can then be compared to images taken from, e.g., the Hubble Space Telescope and new clues about the dark Universe can be obtained, such as the properties of the dark matter that makes up the lensing halo.

Acknowledgment: The authors were supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. This research used resources of the ALCF, which is supported by DOE/SC under contract DE-AC02-06CH11357 and resources of the OLCF, which is supported by DOE/SC under contract DE-AC05-00OR22725.

References

[1] A. Abate et al. [LSST Dark Energy Science Collaboration], arXiv:1211.0310 [astro-ph.CO].

[2] M. Hereld, J.A. Insley, E.C. Olson, M. Papka, V. Vishwanath, M.L. Norman, and R. Wagner, 2011 IEEE Symposium on Large Data Analysis and Visualization (LDAV), 133 (2011).

[3] S. Habib et al., Journal of Physics: Conference Series 180, 012019 (2009).

[4] A. Pope, S. Habib, Z. Lukic, D. Daniel, P. Fasel, N. Desai, and K. Heitmann, Computing in Science and Engineering 12, 17 (2010).

[5] S. Swaminarayan, *Roadrunner: The Dawn of Accelerated Computing*, in Contemporary High Performance Computing, edited by J.S. Vetter (CRC Press, 2013).

[6] For a review of cosmological simulation methods, see K.S. Dolag, S. Borgani, S. Schindler, A. Diaferio, and A.M. Bykov, Space Sci. Rev. 134, 229 (2008).

[7] S. Habib et al., SC12, arXiv:1211.4864

[8] R.W. Hockney and J.W. Eastwood, *Computer Simulation Using Particles* (New York: Adam Hilger, 1988).

[9] S. Habib et al, SC13 proceedings.

[10] J. Woodring, K. Heitmann, J. Ahrens, P. Fasel, C.-H. Hsu, S. Habib and A. Pope, Astrophys. J. Suppl. 195, 11 (2011)

[11] G. Kauffmann A. Nusser, and M. Steinmetz, MNRAS, 286, 795 (1997); Y.P. Jing, H.J. Mo, and G. Borner,~G. 1998, Astrophys. J., 494, 1 (1998); A.J. Benson, S. Cole, C.S. Frenk, C.M. Baugh, and C.G. Lacey, MNRAS, 311, 793 (2000); J.P. Peacock and R.E. Smith., MNRAS, 318, 1144 (2000); U. Seljak, MNRAS, 318, 203 (2000); A.A. Berlind and D.H. Weinberg, Astrophys. J., 575, 587 (2002).

[12] L. Anderson et al. [BOSS Collaboration], arXiv:1312.4877 [astro-ph.CO].

[13] J. Kwan, K. Heitmann, S. Habib, N. Padmanabhan, H. Finkel, N. Frontiere and A. Pope, arXiv:1311.6444 [astro-ph.CO].

[14] A. Benson, New Astronomy, 17, 175 (2012)

Author Information

Katrin Heitmann: Katrin Heitmann is a member of the scientific staff at Argonne National Laboratory in High Energy Physics and Mathematics and Computational Science Divisions. She is also a Senior Fellow at the Computation Institute and the Kavli Institute for Cosmological Physics at the University of Chicago. Her research focuses on physical cosmology, advanced statistical methods, and large scale computing. Heitmann received her PhD in 2000 at the University of Dortmund (Germany), held a postdoctoral position and later a staff position at Los Alamos National Laboratory before she joined Argonne in 2011. She is a member of the American Physical Society.

9700 South Cass Avenue
Argonne National Laboratory
Building 360, Room C-116
Argonne, IL 60439
Phone: (630) 252-1114
Email: heitmann@anl.gov

Salman Habib: Salman Habib is a Senior Scientist at Argonne National Laboratory in High Energy Physics and Mathematics and Computational Science Divisions. He is also a Senior Fellow at the Computation Institute and Senior Member of the Kavli Institute for Cosmological Physics at the University of Chicago. Habib's research covers a broad area in quantum and classical dynamical systems and field theory, including the use of large-scale computing resources to solve problems in these fields. A recent focus of his work has been the application of advanced statistical methods and supercomputing to physical cosmology. Habib received his PhD in 1998 at the University of Maryland, held a postdoctoral position at the University of British Columbia and postdoctoral and staff positions at Los Alamos National Laboratory, before joining Argonne in 2011. He is a member of the American Physical Society.

9700 South Cass Avenue
Argonne National Laboratory
Building 360, Room C-116
Argonne, IL 60439
Phone: (630) 252-1110
Email: habib@anl.gov

Hal Finkel: Hal Finkel is a Computational Scientist at Argonne National Laboratory's Leadership Computing Facility. Finkel's research covers several areas in theoretical cosmology, numerical algorithms and compiler technology, with a focus on applications requiring large-scale computing. Finkel received in PhD in 2011 from Yale University and joined Argonne as a postdoctoral researcher that same year.

9700 South Cass Avenue
Argonne National Laboratory
Building 240, Room 2124
Argonne, IL 60439
Phone: (630) 252-0023
Email: hfinkel@anl.gov

Nicholas Frontiere: Nicholas Frontiere is currently a graduate student at the University of Chicago performing research at Argonne National Laboratory in the High Energy Physics Division. He has received the University of Chicago Nambu Fellowship and the Department of Energy Computational Science Graduate Fellowship. He is currently exploring the application of high performance computing in large-scale cosmology simulations, as well as the scalability of statistical algorithms on such machines. He received a double B.S. in physics and mathematics from UCLA in 2013.

9700 South Cass Avenue
Argonne National Laboratory

Building 360, MS 6
Argonne, IL 60439
Phone: (630) 252-0023
Email: nfrontiere@gmail.com

Adrian Pope: Adrian Pope is currently the Arthur Holly Compton postdoctoral fellow in the High Energy Physics Division at Argonne National Laboratory. He obtained his Bachelors of Science in physics from Carnegie Mellon University in 1999 and his Masters of Arts and PhD in Physics and Astronomy from Johns Hopkins University in 2003 and 2005, respectively. He was a Junior Scientific Researcher at the University of Hawaii and a Director's postdoctoral fellow and a Richard P. Feynman postdoctoral fellow at Los Alamos National Laboratory before joining Argonne. Adrian is interested in cosmological measurements from the statistical clustering of galaxies in astronomical sky surveys and has worked on astronomical instrumentation, data analysis, parameter estimation, statistical methods, and gravitational N-body simulations on high performance computing systems. He has a membership in the American Astronomical Society and was a builder of the Sloan Digital Sky Survey.

High Energy Physics Division
Bldg. 360, MS 6
Argonne National Laboratory
9700 S. Cass Avenue Argonne, IL 60439
Phone: (630) 252-1160
Email: apope@anl.gov

Vitali Morozov: Vitali Morozov is a Principal Application Performance Engineer at the Argonne Leadership Computing Facility. He received his B.S. in Mathematics from Novosibirsk State University, and a Ph.D. in Computer Science from Ershov's Institute for Informatics Systems, Novosibirsk, Russia. At Argonne since 2001, he has been working on computer simulation of plasma generation, plasma material interactions, plasma thermal and optical properties, and applications to laser and discharge-produced plasmas. At the ALCF, he has been working on performance projections and studying the hardware trends in HPC, as well as porting and tuning applications - primarily on Blue Gene supercomputers.

9700 South Cass Avenue
Argonne National Laboratory
Building 240, Room 1127
Argonne, IL 60439
Phone: (630) 252-7068
Email: morozov@anl.gov

Steve Rangel: Steve Rangel is a doctoral student in the Electrical Engineering and Computer Science department at Northwestern University under the supervision of

Dr. Alok Choudhary. He is conducting his thesis research at Argonne National Laboratory in the High Energy Physics division and is co-advised by Salman Habib. Current research interests include large-scale data analysis, high-performance computing, and scalable algorithm design.

Northwestern University
2145 Sheridan Rd. Room LG65
Evanston, IL 60208-3100
Email: steverangel@gmail.com

Eve Kovacs: Eve Kovacs is a scientist/programmer in the HEP division at Argonne National Laboratory. She is a member of the Dark Energy Survey (DES) and works with the DES Supernova group on supernova cosmology and systematics analysis. Kovacs also works on cosmology simulations and semi-analytic galaxy modeling. Prior to working in cosmology, Kovacs was a member of the CDF Collaboration and worked on QCD phenomenology and jet physics. Earlier in her career she worked in theoretical particle physics on a number of topics in lattice gauge theories. She received her PhD from the University of Melbourne in 1980.

HEP, Building 360,
9700 South Cass Avenue
Argonne National Laboratory
Argonne, IL 60439
Phone: (630) 252-6201
Email: kovacs@anl.gov

Juliana Kwan: Juliana Kwan is a postdoc at Argonne National Laboratory in the High Energy Physics Division. She obtained her PhD in Physics from The University of Sydney in 2012. Juliana's research interests include growth of large-scale structure probes and modelling galaxy distributions using N-body simulations for precision cosmology.

HEP, Building 360,
9700 South Cass Avenue
Argonne National Laboratory
Argonne, IL 60439
Phone: (630) 252-1111
Email: jkwan@anl.gov

Nan Li: Nan Li holds a joint postdoctoral position at the University of Chicago and Argonne National Laboratory. His research currently focuses on simulations of strong gravitational lensing. He is interested in gravitational lensing, numerical simulation, and computational cosmology.

9700 South Cass Avenue
Argonne National Laboratory
Building 360, Room C-124
Argonne, IL 60439
Phone: (312) 316-0898
Email: linan7788626@oddjob.uchicago.edu

Silvio Rizzi: Silvio Rizzi is a postdoctoral appointee in data analysis and visualization at the Argonne Leadership Computing Facility. His research interests include large-scale data visualization, augmented reality and immersive environments for scientific research, and computer-based medical simulation. He has an MS in Electrical and Computer Engineering and a PhD in Industrial Engineering and Operations Research from the University of Illinois-Chicago.

9700 South Cass Avenue
Argonne National Laboratory
Building 240, Room 4140
Argonne, IL 60439
Phone: (630) 252-0022
Email: srizzi@anl.gov

Joe Insley: Joseph A. Insley is a principal software development specialist at Argonne National Laboratory and at the University of Chicago. His research interests include the development of parallel and scalable methods for large-scale data analysis and visualization on current and next-generation systems. Insley has an MS in computer science from the University of Illinois at Chicago, and is a senior member of the IEEE Computer Society and of the ACM.

9700 South Cass Avenue
Argonne National Laboratory
Argonne, IL 60439
Phone: (630) 252-5649
Email: insley@anl.gov

Venkat Vishwanath: Venkatram Vishwanath is an Assistant Computer Scientist in the Mathematics and Computer Science Division at Argonne National Laboratory. He is also a member of the Argonne Leadership Computing Facility. Dr. Vishwanath joined Argonne as an Argonne Scholar and Argonne Director's Fellow in 2009. His areas of research include runtime and programming models for data-intensive computing, scalable algorithms for data movement, scientific data visualization, and performance analysis for parallel applications. He completed his doctorate degree in computer science in 2009 from the University of Illinois at Chicago. He is also a Fellow of the Computation Institute at the University of Chicago and an adjunct professor in the department of computer science at the Northern Illinois University.

9700 South Cass Avenue
Argonne National Laboratory
Building 240, Room 4141
Argonne, IL 60439
Phone: (630)-252-4971
Email: venkat@anl.gov

Tom Peterka: Tom Peterka is an assistant computer scientist at Argonne National Laboratory, fellow at the Computation Institute of the University of Chicago, adjunct assistant professor at the University of Illinois at Chicago, and member of the IEEE. His research interests are in large-scale parallelism for in situ analysis of scientific datasets. His work has led to two best paper awards and publications in ACM SIGGRAPH, IEEE VR, IEEE TVCG, and ACM/IEEE SC, among others. Tom received his Ph.D. in computer science from the University of Illinois at Chicago in 2007.

9700 South Cass Avenue
Argonne National Laboratory
Building 240, Room 3143
Argonne, IL 60439
Phone: (630) 252-7198
Email: tpeterka@mcs.anl.gov

David Daniel: David Daniel is a scientist in the Applied Computer Science group at Los Alamos National Laboratory, and has worked on a wide variety of problems in scientific computing and scalable system software, including cosmology, lattice QCD, and multi-physics simulation codes. He has made major contributions to MPI libraries including Open MPI and LA-MPI. Dr. Daniel has a B. Sc. in physics from Imperial College, London, and a Ph. D. in theoretical physics from the University of Edinburgh.

Los Alamos National Laboratory
CCS-1, CCS Division
Los Alamos, NM 87545
Phone: (505) 665-4939
Email: ddd@anl.gov

Patricia Fasel: Patricia Fasel is a Scientist III in the Information Sciences group at Los Alamos National Laboratory. She has degrees in computer science and mathematics from Purdue University and works in algorithm development, software engineering and data analysis in support of scientific applications. Her interests include in situ analysis and visualization of large scale simulations, parallel programming, and large scale data mining.

Los Alamos National Laboratory
CCS-3, MS B265, CCS Division

Los Alamos, NM 87545
Email: pkf@anl.gov

George Zagaris: George Zagaris is an R&D Engineer in the Scientific Computing division at Kitware Inc., where his current focus is the development of a framework for in situ analysis of large-scale cosmological simulations. Zagaris received his MS (2008) in Computer Science from the College of William & Mary. His research interests and expertise lie at the intersection of computational science and computer science in the context of High Performance Computing.

Kitware, Inc.
28 Corporate Drive
Clifton Park, NY 12065
Email: george.zagaris@kitware.com